Homogeneity and identity tests for unidimensional Poisson processes with an application to neurophysiological peri-stimulus time histograms

Christophe Pouzat, Antoine Chaffiol, Avner Bar-Hen

Homogeneity and identity tests for unidimensional Poisson processes with an application to neurophysiological peri-stimulus time histograms.

Christophe Pouzat[1],
Antoine Chaffiol[2],
Avner Bar-Hen[1]

[1] MAP5, Paris-Descartes University and CNRS UMR 8145

[2] Pierre and Marie Curie University, CNRS UMR 7210 and INSERM UMR U968

November 4, 2015

Abstract

The first characterization of the response of a neuron to a stimulus is the peri-stimulus time histogram (PSTH). From a statistical viewpoint the PSTH is an estimator of the intensity of the inhomogeneous Poisson process describing (asymptotically) the aggregated responses of the neuron to repeated presentations of the stimulus. The PSTH is often used to address *qualitatively* two questions: i) is the neuron responding to the stimulation? ii) are the responses of a neuron to two different stimuli different? We propose here *quantitative* answers based on the PSTH. The observed state space is first finely binned before applying a variance stabilizing transformation. The homogeneity ("Is the neuron responding?") is then addressed by using a linear smoother estimator for the scaled Poisson process intensity before building a confidence set containing this estimator. The identity ("Are the two responses identical?") is addressed by constructing the cumulative sum of the difference of the scaled PSTH obtained with the two stimuli. This cumulative sum tends under the null hypothesis towards a canonical Brownian motion process. Minimal surface domains containing the totality of a given fraction of the realizations of a canonical Brownian motion process are available, allowing us to build an identity test. This identity test can also be used to compare the neuron's activity before and after the stimulus presentation making also an homogeneity test. Our motivating dataset arises from our own experimental work and is publicly available. Our proposed methods are implemented in publicly available and documented codes for either R or Python.

1 Introduction

Neurophysiologists often repeat the presentation of a given stimulus, *e.g.*, an odor when studying the olfactory system, and then, observe neuronal outputs

(sequences / trains of action potentials / spikes). The outputs are time aligned on the stimulus onset and their sum is generally modeled as a unidimensional Poisson process since a key result from the theory of point processes states that if the successive responses are independent, the resulting process converges towards an *inhomogeneous Poisson process* (Grigelionis, 1963; Brown, 1978; Ventura et al., 2002). This paper focuses on two classical questions in neurophysiology: i) is the process homogeneous (is the neuron responding to the stimulation)? ii) are two given observations identical (are the responses to two different stimuli different)?

A classical approach is to build the peri-stimulus time histogram or `PSTH` (Gerstein and Kiang, 1960; Perkel et al., 1967), an estimator of the intensity of the Poisson process describing the aggregation of the successive responses. The PSTH has been around for a long time and remains the first descriptive statistics used when characterizing stimulus / response relations of neurons. Using classical statistical tools, we provide a framework to use the PSTH as a quantitative tool. The interest of deriving distributional properties of the PSTH is twofold since it leads to quantitative answers to our two questions: (i) is the Poisson process homogeneous, *i.e.* is the intensity constant along the signal— a new test of homogeneity is proposed in Section 4—; (ii) are two PSTH identical—an identity test is proposed in Section 5—.

Homogeneity test have been proposed in the literature (Cox and Lewis, 1966) and are reviewed in section 2. The assessment of differences between two PSTH has received some attention (Ellaway, 1978; Dörrscheidt, 1981; Ushiba et al., 2002; Tam, 2009) going generally through an examination of the cumulative difference of the PSTH and linking this question to the interpretation of cumulative sum charts (Siegmund, 1985; Hawkins and Olwell, 1998). The quantitative evaluation of the difference is generally done through pointwise intervals; that can be appropriate for change point detection but fall short of addressing the overall difference issue. When constructing the PSTH, the automatic bin width choice, or more generally the bandwidth choice when working with smooth estimates, has rarely been addressed despite of its importance (Kaufman et al., 2005; Shimazaki and Shinomoto, 2007; Wallstrom et al., 2007; Pouzat and Chaffiol, 2009; Shimazaki and Shinomoto, 2010; Reynaud-Bouret et al., 2014).

Following Cox and Lewis (1966, Sec. 3.2, pp 43-44) we finely bin the observed state space before applying a *variance stabilizing transformation* (Anscombe, 1948; Freeman and Tukey, 1950; Brown et al., 2010). The homogeneity is then first addressed by using a *linear smoother* (Wasserman, 2006) estimator for the scaled Poisson process intensity before building a *confidence set* containing this estimator (Sun and Loader, 1994). We are therefore *not performing kernel density estimation* (KDE). KDE does not require prior binning of the data but does not lead to the construction of confidence sets. The prior variance stabilization is important since we justify the construction of the confidence sets by using Gaussian processes and this requires the observation errors to be IID and to follow a Gaussian distribution. In a second time, we address the question of identity of two observed processes by applying a common binning on two observations followed by variance stabilization and subtraction, bin per bin of the two scaled processes. The resulting sequence of differences should then be, under the null hypothesis of identity, a sequence of IID draws from a *centered* Gaussian distribution with a known variance. We take into account multiple comparisons by first constructing the *cumulative process* (the differences con-

tained in the successive bins are added one by one in the bin order) that should under the null hypothesis—and after proper rescaling—converge towards a standard Brownian motion process as stated in Donsker's theorem (Billingsley, 1999; Durrett, 2009). The question then reduces to testing if the rescaled cumsum of the difference process is compatible with a standard Brownian motion process. We use the results of Kendall et al. (2007) to build a minimal surface domain within which a prescribed fraction of the observed standard Brownian motion processs will be *entirely* contained. The combination of Donsker's theorem with the minimal surface domain allows us to built a test. This test is free since it does not depend on the true common underlying intensity of the two observed process under the null hypothesis. This test can also be used to compare the neuron's activity before and after stimulus presentation—when enough data are available—giving a second way of addressing the homogeneity issue. The proposed tests are implemented in two software environments: `R` and `Python`. The technique used in the second test can clearly be applied to any context where a sufficient number (50 or more) of tests are performed.

Section 3 presents briefly the neuronal data used in this article and the kernel based linear smoother used in section 4. A simulation study of the sample size sensitivity of the identity test is presented in Section 6. In Section 7, a discussion of the proposed test is carried out and perspectives are presented. The Supplementary Files allow readers both to reproduce the complete analysis presented in this article and to apply it to the complete data set (or a new data set) using `R` or `Python`.

2 Existing homogeneity tests

Cox and Lewis (1966) present tests for homogeneous Poisson (Sec. 6.3) and renewal (Sec. 6.4) processes. The tests for Poisson processes use the fact that if the observed times: $\{t_1, t_2, \ldots, t_n\}$ are a realization of a homogeneous Poisson process with rate λ on the time interval $[0, t_0]$, then, conditionally on n, the total number of events observed at the end of the time period, the quantities: $\{u_{(i)} = t_i/t_0\}_{i=1,\ldots,n}$ are observations of the order statistics of n IID draws from a uniform distribution on $(0, 1)$. It is then possible to apply a Kolmogorov test or an Anderson-Darling test against this null hypothesis giving a *uniform conditional test for a Poisson process*. Durbin (1961, p. 48) followed by Lewis (1965) argue further for the use of what Cox and Lewis (1966, pp 154-155) dubbed *Durbin's transformation* of the t_i in order to improve the power of these tests against the uniform null hypothesis. The algorithm producing this transformation follows:

1. Go from the $\{u_{(i)} = t_i/t_0\}_{i=1,\ldots,n}$ discussed in the previous paragraph to the intervals: $\{c_1 = u_{(1)}, c_i = u_{(i)} - u_{(i-1)}\ (i = 2, \ldots, n), c_{n+1} = 1 - u_{(n)}\}$ (the latter should IID realizations from an exponential distribution with parameter 1).

2. Get the order statistics $\{c_{(1)}, \ldots, c_{(n)}\}$ and form the differences $g_i = (n + 2 - i)\left(c_{(i)} - c_{(i-1)}\right)$ for $i = 1, \ldots, n + 1$ with $c_{(0)} = 0$ (they should be independent exponentially distributed random variables with means 1).

3. The observations $u'_{(i)} = \sum_{j=1}^{i} g_j$ for $i = 1, \ldots, n$ should then be observa-

tions from the order statistics of n IID draws from a uniform distribution on $(0, 1)$.

As pointed out by Cox and Lewis (1966, p. 158) the tests on transformed data are sensitive to discretization: they fail to apply if the latter is too coarse. The data used in this manuscript were sampled at 12800 Hz with a spike sorting procedure (Sec. 3.1) that did not properly cope with sampling jitter (Pouzat and Detorakis, 2014). This unaccounted for sampling jitter amounts to a "too coarse" sampling and give rise to a pronounced stair-case aspect of the empirical cumulative distribution function (ECDF) of the $u'_{(i)}$ for small values of i. This leads to spurious positive values when applying the Anderson-Darling test. We therefore decided when working with the transformed data to jitter the original observed times uniformly by plus or minus half a sampling period (in practice plus or minus 40 μs). This destroys the stair-case aspect without touching the overall structure as illustrated and investigated in our Supplementary Files.

In addition to these tests against a uniform distribution on $(0, 1)$, the correlation coefficients of the successive inter-event intervals at different lags (the autocorrelation function of the inter-events intervals) should be inspected and the log of the survivors function—that should be a straight line under the null hypothesis—should be plotted.

3 Peri-stimulus time histograms and their smooth estimates

3.1 The neurophysiological data and the two questions

Our datasets arise from our own experimental work. They are publicly available either as part of our `R` package `STAR` (on the CRAN) or from the zenodo server (`DOI:10.5281/zenodo.14281`, Pouzat and Chaffiol 2015). The data are fully described in Pouzat and Chaffiol (2009). After recording the spontaneous activity for a few minutes, cockroaches (*Periplaneta americana*) were stimulated with odors (citronellal, terpineol, etc). The recordings were made from the first olfactory relay of the insect with an interval between stimulations long enough for the effect of the former to have disappeared when the next comes. After a preprocessing called *spike sorting* (Eggermont, 1990; Einevoll et al., 2012), sequences of action potentials or spike trains were extracted from the continuous raw data. We then got data representations called *raster plots* (Fig. 5, in the Appendix, Brillinger 1992). A very common question is then: is the neuron responding? A second question of interest is to decide whether the responses for the same neuron stimulated with different odors are identical or not (Fig. 6, in the Appendix). This type of stimulus-response study is very common in neurophysiology where experimentailists are trying to characterize the response spectrum of neurons. For the data used here this means that many odors were systematically presented to the animal and, for each neuron recorded in the first olfactory relay, the number of odors generating a response (first question above) was looked for and, when a neuron did respond to several odors, the responses identity (second question above) was addressed. Ultimately, from a behavioral viewpoint, neurophysiologists try to understand how animals (like us or like insects) can identify a huge number of odors with a small ($<$ 100-1000) number

of different olfactory receptor types.

3.2 The "classical" peri-stimulus time histogram

As mentioned in the introduction, the aggregated process ("summation" of all the responses) converges, under smooth assumptions, towards an inhomogeneous Poisson process. A peri-stimulus time histogram (PSTH) (Gerstein and Kiang, 1960; Perkel et al., 1967), is, in statistical terminology, a piecewise constant estimator of the Poisson process intensity, $\lambda(t)$, built conditionally on a bin width of duration δ. The number of spikes falling in each of the k bins (from all the trials realigned on the stimulus onset) is counted and noted $\{y_1, \ldots, y_k\}$. These observations are realizations from a set of Poisson random variables $\{Y_1, \ldots, Y_k\}$ with parameters:

$$n \int_{t_i-\delta/2}^{t_i+\delta/2} \lambda(u)\, du \;\approx\; n\,\lambda(t_i)\,\delta\,, \quad i = 1, \ldots, k\,, \tag{1}$$

where t_i is the center of a class (bin) and n is the number of stimulations. The piecewise constant estimator of $\lambda(t)$ is then defined by:

$$\hat{\lambda}(t) = y_i/(n\delta)\,, \quad \text{if} \quad t \in [t_i - \delta/2, t_i + \delta/2)\,. \tag{2}$$

Our two motivating questions can be stated as statistical tests:

- Homogeneity test correspond to the test of $H_0 : \lambda(t) = \lambda$ versus $H_1 : \exists t_1$ and t_2 such that $\lambda(t_1) \neq \lambda(t_2)$.
- Identity test: Let $\lambda_1(t)$ and $\lambda_2(t)$ be the intensities of two Poisson process derived from the PSTH of a single neuron from two different stimulations, then we test $H_0 : \lambda_1(t) = \lambda_2(t)$ for all t versus $H_1 : \exists t$ such that $\lambda_1(t) \neq \lambda_2(t)$.

3.3 Variance stabilization before smooth estimation

Our goal here is not intensity estimation per se, but the construction of a confidence set containing the actual $\lambda(t)$ (for all t), with a given probability. We first transform the Poisson regression problem to a Gaussian regression by stabilizing the variance (Anscombe, 1948; Freeman and Tukey, 1950; Brown et al., 2010) with a square root transform.

Let $X_{i,j}$ be the number of spikes falling in bin i for trial j and let $Y_i = \sum_{j=1}^{n} X_{i,j}$ be the aggregated process. As noted by Freeman and Tukey (1950), the best overall stabilization for small counts if obtained by transforming Y_i as:

$$Z_i = \sqrt{Y_i} + \sqrt{Y_i + 1}\,. \tag{3}$$

The variance of Z_i is then 1. Our software implementations consider the transformation as a parameter, the user can work with Freeman-Tukey transformation as well as Anscombe transformation ($Z_i = 2\sqrt{Y_i + 3/8}$) or the one of Brown et al. ($Z_i = 2\sqrt{Y_i + 1/4}$).

The bin width should be chosen large enough to have a few events per bin most of the time. We set this width such that an expected number of 3 events

per bin was obtained using the spontaneous frequency estimated from 60 seconds long recordings without stimulus presentation—more specifically, the width was set to the smallest millisecond larger or equal to the targeted count divided by the product of the spontaneous frequency and the number of trials. Following Freeman and Tukey (1950), an expected number of 3 events is sufficient to ensure the stability of the transformed variance.

3.4 Smooth estimation: Nadaraya-Watson estimator

After variance stabilization, we have a Gaussian regression setting:

$$Z_i = r(t_i) + \epsilon_i \,, \tag{4}$$

where the $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Following Wasserman (2006) we use a Nadaraya-Watson estimator:

$$\hat{r}(t) = \sum_{i=1}^{k} l_i(t) Z_i \,.$$

The functions l_i are defined by:

$$l_i(t) = \frac{K\left(\frac{t-t_i}{h}\right)}{\sum_{j=1}^{k} K\left(\frac{t-t_j}{h}\right)} \,. \tag{5}$$

In this article we use the tricube kernel: $K(t) = 70/81 \left(1 - |t|^3\right)^3 \mathbb{I}(t)$, where $\mathbb{I}(t)$ is the indicator function of interval $[-1, 1]$.

Since after variance stabilization the variance is known we can set our bandwidth by minimizing Mallows' C_p criterion instead of using cross-validation. More explicitly, we minimize:

$$(1/k) \sum_{i=1}^{k} \left(Z_i - \hat{r}(t_i)\right)^2 + 2 \left(\sum_{i=1}^{k} l_i(t_i) \right) / k \,.$$

Fig. 1 illustrates this procedure with the citronnellal response of neuron 2 from data set `e070528`. Note that the optimal bandwidth (110 ms) is ten times larger than the initial bin width (11 ms).

4 Confidence set and homogeneity test

Let $\mathcal{S}$ be a large class of functions, we would like to provide a *confidence envelop* $\mathcal{B} = \{s \in \mathcal{S} : u_1(t) \leq s(t) \leq u_2(t),\ \forall t \in [a,b]\}$, such that:

$$\Pr\{r \in \mathcal{B}\} \geq 1 - \alpha \tag{6}$$

for all $r \in \mathcal{S}$ (Wasserman, 2006). Since smooth estimators exhibit a bias that does not disappear even with large sample sizes, we built sets around $\overline{r} = \mathrm{E}(\hat{r}) = \sum_{i=1}^{k} l_i(t) r(t_i)$. Using the fact that the estimator is a linear function of the observations and that the latter follow (asymptotically) a Gaussian distribution centered on the true value with a unit variance we have:

$$\mathrm{Var}\left(\hat{r}(t)\right) = \sum_{i=1}^{k} l_i(t)^2 \doteq \|l(t)\|^2 \,.$$

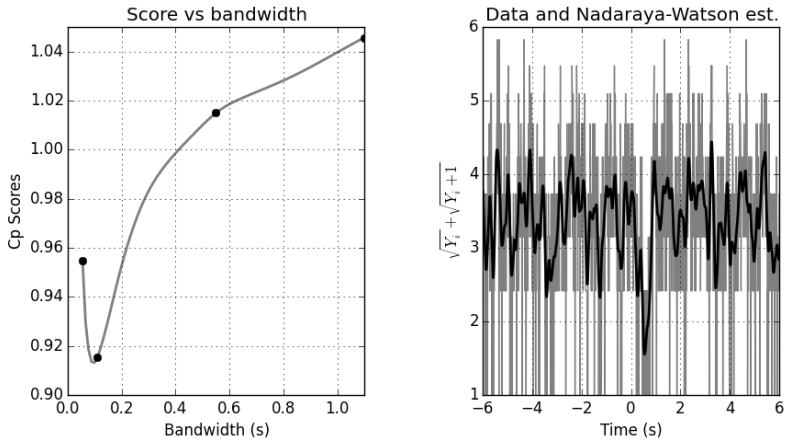

Figure 1: Nadaraya-Watson estimator of Neuron 2 response to citronellal (from experiment `e070528`). Left: Mallows' Cp score vs bandwidth. The dots correspond to the "discrete set" of bandwiths chosen *before seeing the data* to build the confidence bands (see section 4). The line corresponds to a "continuous set". Right: variance stabilized spike counts per bin (grey) and Nadaraya-Watson estimator with optimal bandwidth from the discrete set (black). The bin width used to build the stabilized PSTH (grey) was 11 ms long, the optimal one in the discrete set is ten times as large. The stimulus onset comes at time 0.

We consider a confidence band for $\overline{r}(t)$ of the form:

$$I(t) = (\hat{r}(t) - c\|l(t)\|, \hat{r}(t) + c\|l(t)\|) \ ,$$

with $c > 0$ and $a \leq t \leq b$.

Following Sun and Loader (1994) and Wasserman we have:

$$\Pr\left\{\overline{r}(t) \notin I(t) \text{ for some } t \in [a,b]\right\} = \left\{ \begin{array}{l} \Pr\left\{\max_{t\in[a,b]} \frac{|\hat{r}(t)-\overline{r}(t)|}{\|l(t)\|} > c\right\} \ , \\ \Pr\left\{\max_{t\in[a,b]} \frac{|\sum_{i=1}^{k} \epsilon_i l_i(t)|}{\|l(t)\|} > c\right\} \ , \\ \Pr\left\{\max_{t\in[a,b]} |W(t)| > c\right\} \ , \end{array} \right.$$

where $W(t) \doteq \sum_{i=1}^{k} \epsilon_i l_i(t)/\|l(t)\|$ is a *Gaussian process*.

The constant c is given by the solution the *tube formula* (Sun and Loader) for the distribution of the maximum of a Gaussian process:

$$\alpha = \Pr\left\{\max_{t\in[a,b]} |\sum_{i=1}^{k} \epsilon_i l_i(t)/\|l(t)\|| > c\right\} \approx 2\left(1 - \Phi(c)\right) + \frac{\kappa_0}{\pi} \exp -\frac{c^2}{2} \ ,$$

and $\kappa_0 \approx (b-a)/h \left(\int_a^b K'(t)^2 dt\right)^{1/2}$. This confidence envelop / band construction assumes implicitly that the kernel if fixed, that is the kernel bandwidth (parameter h in Eq. 5) is known before observing the data. Although we will argue in the discussion that we can reasonably choose this bandwidth (in the 500 ms — 1 s range) in the context of our system, we present here a more general approach following the development of Wasserman. We chose a discrete set

of 5 bandwidths—the bin width multiplied by 5, 10, 50, 100 and 500; the black dots on Fig. 1, left; the rightmost dot is not shown—from which the best was chosen by minimizing the C_p. In the context of testing this consideration of 5 different models (one for each bandwidth) amounts to multiple comparison for which a Bonferroni correction was applied. Fig. 2 illustrates the construction of a 95 % confidence envelop for the Nadaraya-Watson estimator of the citronellal response of Neuron 2. If the underlying intensity was homogeneous, the (transformed) intensity should be a constant and a horizontal line like the black line of Fig. 2 should be *entirely* contained within the envelop. Based on Fig. 2 we can reject the null hypothesis of homogeneity at the 95 % level. This conclusion can clearly be reached without contructing the figure since the null hypothesis can be accepted as long as the maximum of the lower boundary is lower than the minimum of the upper one.

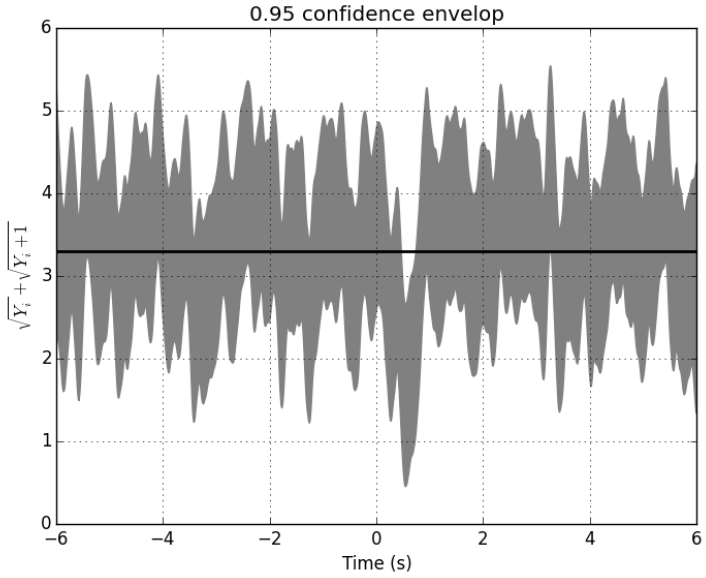

Figure 2: Confidence envelop at 0.95 level for the Nadaraya-Watson estimator of Neuron 2 response to citronellal. In black, an horizontal line completely within the band prior to the stimulus but not after it. The stimulus onset is at time 0.

5 Identity test

We now turn to our second question: the identity of the responses of a given neuron to two different stimuli; and look for an answer that does not require any hypothesis on the underlying process intensity, when the latter is actually the same for the two situations. We could clearly use our previous confidence set construction, setting the guaranteed probability of containing the actual smooth intensity through a Bonferroni correction, such that the two bands would not overlap at, at least, one time point with a given frequency under the null

hypothesis of identity. But we can find a more direct answer that is also, in principle, better since the inequality in Eq. (6) becomes (asymptotically) an equality.

We nevertheless start as before by using a sequence of Poisson random variables as a proxy for our process through time discretization. So when we compare the citronellal and terpineol responses of Neuron 1 (from experiment `e060817`, Fig. 6, in the Appendix), the bin width is automatically set at 18 ms and we get two sets of observations: $\{y_1^{citron}, \ldots, y_k^{citron}\}$ and $\{y_1^{terpi}, \ldots, y_k^{terpi}\}$. We also stabilize the variance as we did before ($z_i = \sqrt{y_i} + \sqrt{y_i + 1}$) to get: $\{z_1^{citron}, \ldots, z_k^{citron}\}$ and $\{z_1^{terpi}, \ldots, z_k^{terpi}\}$. Our null hypothesis is that the two underlying inhomogeneous Poisson processes are the same, that is:

$$z_i^{citron} = r(t_i) + \epsilon_i^{citron} \quad \text{and} \quad z_i^{terpi} = r(t_i) + \epsilon_i^{terpi}\,,$$

leading us to:

$$\frac{Z_i^{terpi} - Z_i^{citron}}{\sqrt{2}} = \epsilon_i\,. \tag{7}$$

We therefore want to test the null hypothesis H_0 : the collection of observed differences $\frac{Z_i^{terpi} - Z_i^{citron}}{\sqrt{2}} \sim \mathcal{N}(0,1)$. Under H_0 we can apply Donsker theorem (Billingsley, 1999; Durrett, 2009) that states that the sequence of processes:

$$S_k(t) = \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor kt \rfloor} \frac{Z_i^{terpi} - Z_i^{citron}}{\sqrt{2}}, \quad 0 \le t \le 1, \tag{8}$$

converges in law towards a canonical Brownian motion process.

It is direct to construct $S_k(t)$ and check if the observed trajectory is, or is not, consistent with a Brownian motion process. Ideally, we would like to define a domain in $[0,1] \times \mathbb{R}$ containing the realizations of a canonical Brownian motion process with a given probability. To have a reasonable power, we would like the surface of this domain to be minimal, but Kendall et al. (2007) showed that the upper boundary of this minimal surface domain is given by:

$$u^*(t) \equiv \sqrt{-W_{-1}\left(-(\kappa t)^2\right)}\,\sqrt{t}, \quad \text{for} \quad \kappa\, t \le 1/\sqrt{e}$$

where W_{-1} is the secondary real branch of the Lambert W function (defined as the solution of $W(z) \exp W(z) = z$); κ being adjusted to get the desired probability. An almost minimal surface that is simpler to work with has its upper boundary given by: $u(t) = a + b\sqrt{t}$ (Kendall et al., 2007). An efficient and simple algorithm for adjusting a and b or κ is found in Loader and Deely (1987). The comparison between the citronellal and terpineol responses of Neuron 1 (from experiment `e060817`) is shown on Fig. 3. Fig. 7 in the Appendix, showing the smooth stabilized PSTH of the two responses, helps understanding the observed motion (black on Fig. 3). The peak response of terpineol is larger than the one of citronellal accounting for the sharp upward jump before 0.5 normalized time on Fig. 3. Then, between second 1.5 and second 6, the terpineol response remains slightly above the citronellal one (Fig. 7) explaining the constant upward drift after normalized time 0.5 on Fig. 3. The comparison of even and odd stimulus number for terpineol has been added to Fig. 3 as a control since we do not expect here any difference in the underlying intensity.

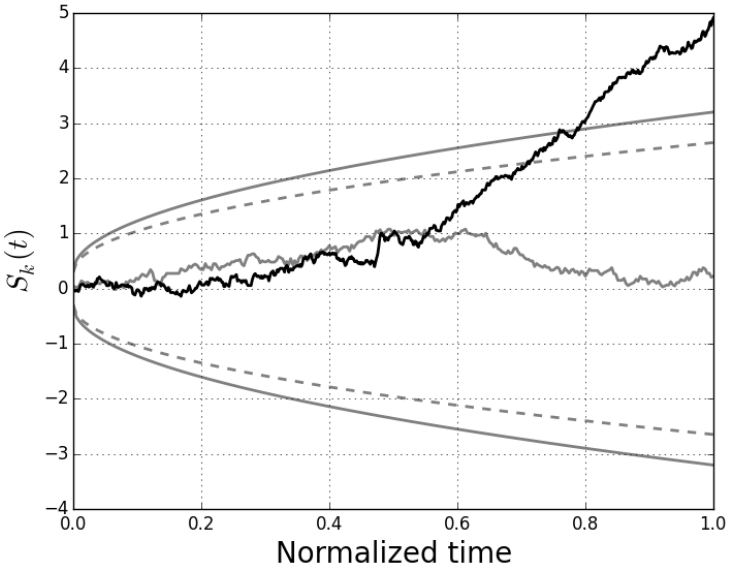

Figure 3: Comparison between the citronellal and terpineol responses of Neuron 1 (from experiment `e060817`). Almost minimal surface domains with probabilities 0.95 (dashed grey) and 0.99 (grey) of containing an observed canonical Brownian motion process. Black: terpineol - citronellal (curve built out of 612 $(z_i^{terpi} - z_i^{citron})/\sqrt{2}$ observations); noisy grey curve: even terpineol trials - odd terpineol trials (curve build out of 315 $(z_i^{terpi,even} - z_i^{terpi,odd})/\sqrt{2}$ observations).

5.1 An additional homogeneity test

The homogeneity question can also be formulated as an identity question: is the process before stimulus onset identical to the one after onset? This requires to have the stimulus onset late enough in the acquisition epoch, or a spontaneous discharge rate large enough, giving a sufficiently large value of parameter k in Eq. 8—since k equals the duration divided by the bin width that is itself inversely proportional to the spontaneous rate. If such is the case, like for the citronellal responses of neuron 2 of experiment `e070528` we considered before (the stimulus onset comes 6 seconds after the beginning of the acquisition epoch and the spontaneous rate is high at 19.6 Hz), we can obtain a sequence (z_i^{before}) with (stabilized) counts before stimulus onset and a corresponding sequence (z_i^{after}) using data recorded after the stimulation. Applying the procedure we just described for the different responses of neuron 1 from experiment `e060817`, we get here Fig. 4. The conclusion is the same as the one drawn from Fig. 2: the neuron is responding to the citronellal stimulation. The portion of the observed motion leaving the confidence region in the upper direction between 0.1 and 0.2 normalized time units corresponds to the dip following the stimulus onset on Fig. 2 (between time 0.5 and 1.0).

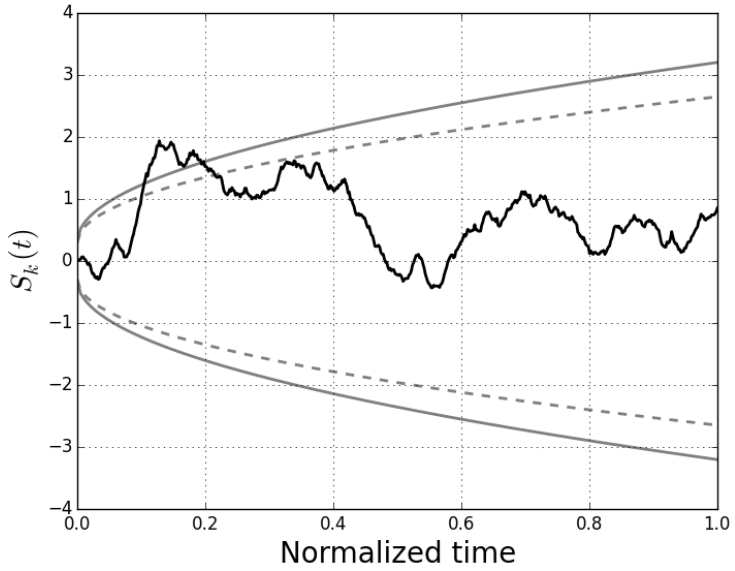

Figure 4: Citronellal response of neuron 2 from experiment `e070528`. The six seconds prior to the stimulus presentation are compared to the 6 seconds after it. Almost minimal surface domains with probabilities 0.95 (dashed grey) and 0.99 (grey) of containing an observed canonical Brownian motion process. The black curve is built out of 546 values of $(z_i^{before} - z_i^{after})/\sqrt{2}$.

6 Sample size sensitivity of the Brownian motion process based test

We estimated the coefficients a and b of the square root boundary for various values of the coverage probability, that is probability to have a realization of a canonical Brownian motion process entirely inside the domain (bounded by the square root boundary). We used the numerical method of Loader and Deely (1987) that also yields an absolute bound on the probability. We thus obtained the results of Table 1.

We then performed a Monte Carlo simulation drawing for each sample size, 100000 replicates. For each replicate, the approximate Brownian motion process was obtained by constructing its cumulative sum and rescaling it. The number of replicates crossing the 10 domains defined by the 10 sets of coefficients of Table 1 was computed. The results are reported as "Agresti-Coull" 95% confidence intervals (Brown et al., 2001) for the empirical coverage probability in Table 2. We can see that the empirical coverage probability approaches systematically its nominal (asymptotic) value from above. We can use this table to correct for the sample size: if a sample size of size 50 is considered and if a coverage probability of 0.95 is requested, the coefficients giving an asymptotic coverage probability of 0.93 should be used.

Table 1: Coefficient a and b for (upper) boundaries given by $a + b\sqrt{t}$ ensuring a given coverage probability. Values have been rounded to the third digit giving enough precision to reproduce the results of Table 2.

Cov. Prob.	a	b
0.99	0.312	2.891
0.98	0.308	2.668
0.97	0.305	2.531
0.96	0.302	2.429
0.95	0.300	2.348
0.94	0.298	2.279
0.93	0.296	2.220
0.92	0.295	2.167
0.91	0.293	2.120
0.90	0.292	2.077

Table 2: Limits of the "Agresti-Coull" 95% confidence intervals of the empirical coverage probability for various sample sizes (horizontal) and various nominal coverage probabilities (vertical). The third decimal of the intervals limits have been rounded upward, respectively downward for the upper, respectively lower, limit.

	25	50	75	100	250	500	750	1000	2500	5000	7500	10000
0.99 up	0.995	0.994	0.994	0.993	0.993	0.993	0.992	0.992	0.992	0.991	0.992	0.991
0.99 low	0.993	0.992	0.991	0.991	0.99	0.991	0.989	0.99	0.989	0.989	0.989	0.989
0.98 up	0.989	0.987	0.986	0.986	0.984	0.984	0.983	0.982	0.982	0.982	0.982	0.981
0.98 low	0.987	0.985	0.984	0.983	0.981	0.982	0.98	0.98	0.979	0.979	0.979	0.978
0.97 up	0.983	0.98	0.979	0.978	0.976	0.975	0.974	0.973	0.972	0.972	0.972	0.972
0.97 low	0.98	0.977	0.976	0.975	0.973	0.972	0.971	0.97	0.969	0.969	0.969	0.969
0.96 up	0.977	0.973	0.971	0.97	0.967	0.966	0.964	0.964	0.963	0.962	0.962	0.962
0.96 low	0.974	0.97	0.968	0.967	0.964	0.962	0.961	0.961	0.959	0.959	0.959	0.959
0.95 up	0.97	0.966	0.964	0.962	0.959	0.957	0.955	0.955	0.954	0.953	0.953	0.952
0.95 low	0.967	0.963	0.961	0.959	0.956	0.954	0.951	0.951	0.95	0.95	0.949	0.948
0.94 up	0.964	0.959	0.956	0.954	0.951	0.948	0.946	0.945	0.944	0.944	0.943	0.943
0.94 low	0.96	0.955	0.952	0.951	0.947	0.944	0.942	0.941	0.94	0.94	0.939	0.939
0.93 up	0.958	0.952	0.948	0.946	0.942	0.939	0.937	0.937	0.934	0.935	0.934	0.933
0.93 low	0.954	0.948	0.944	0.942	0.938	0.935	0.933	0.933	0.93	0.931	0.929	0.929
0.92 up	0.951	0.944	0.94	0.938	0.934	0.93	0.928	0.928	0.925	0.925	0.924	0.923
0.92 low	0.947	0.94	0.936	0.933	0.929	0.925	0.923	0.923	0.92	0.921	0.92	0.919
0.91 up	0.944	0.937	0.932	0.929	0.925	0.921	0.918	0.919	0.916	0.915	0.914	0.913
0.91 low	0.941	0.933	0.927	0.925	0.921	0.916	0.914	0.914	0.911	0.911	0.91	0.909
0.90 up	0.938	0.929	0.923	0.921	0.916	0.912	0.909	0.909	0.906	0.906	0.905	0.904
0.90 low	0.934	0.925	0.919	0.917	0.912	0.907	0.904	0.904	0.901	0.901	0.90	0.899

7 Discussion

Although PSTH have been heavily used for more than 50 years (Wall, 1959; Gerstein and Kiang, 1960) and despite of serious attempts at building a quantitative tool out of an essentially descriptive one (Ellaway, 1978; Dörrscheidt, 1981; Ushiba et al., 2002), this article contains, to our knowledge, the first construction of *PSTH based* quantitative tests for two very common questions in neurophysiology:

1. Is a given neuron responding "on average" to a given stimulus?

2. Are the responses of a given neuron to two different stimuli the same?

We emphasized "PSTH based" since Cox and Lewis (1966) proposed homogeneity tests for Poisson processes. As should be clear from section 2, their tests have the advantage of working on the raw data while our requires prior binning; but

their tests are global while our is local. Our test has the additional advantage of extending a statistic the neurophysiologists are used to work with. We show in the Supplementary files that, when applied to the citronellal response of neuron 2 from data set `e070528`—the one illustrated in section 4—the Anderson-Darling test after Durbin's transformation of the data is the only significant one at the 95% level.

Essential for the validity of our tests is the convergence of the "aggregated responses" making the PSTH towards an inhomogeneous Poisson process. This is a well established result (Grigelionis, 1963; Brown, 1978; Ventura et al., 2002), but practitioners still have to check that enough stimuli have been used for the convergence to have been reached (the Supplementary Files contain a comprehensive description of how this was done on the data analyzed here). Our methods require a discretization of the observed (aggregated) process bringing the problem into an easier to deal with Poisson regression framework. The bin width setting, giving 3 or more events in each of the bins making the PSTH, is not a limitation of our procedure. Experimentalists can in practice always record enough data—in the "spontaneous regime"—to get an estimate of the neurons basal firing rate necessary to set the bin width. After discretization, a well known variance stabilizing transformation (Freeman and Tukey, 1950; Anscombe, 1948; Brown et al., 2010) makes the answer to the first question above a direct application of (advanced) textbook methodology (Wasserman, 2006) through the construction of a confidence set / envelop. It should nevertheless be noted that we are not the first to use variance stabilizing transformation in a neuronal spike train analysis context: Brillinger and colleagues did it in 1976 (Brillinger et al., 1976). The confidence envelop can also be used to estimate the *response delay*, a parameter of interest in neurophysiology: once a response threshold has been set, it is enough to find the two times at which the two boundaries of the confidence envelop cross that threshold in the upward direction.

The homogeneity test illustration we have given in Fig. 2, assumed that little was known *a priori* on the response characteristic time. This led us to consider a (small) set of kernel bandwidths requiring a Bonferroni correction and leading to a larger confidence envelop. But prior knowledge on this system tells us that typical odor responses exhibit a characteristic time of 500 ms to 1 s. In such cases we would therefore set the kernel bandwith to 750 ms and build the confidence envelop directly. Our proposed homogeneity test can clearly be used in both of these contexts.

We address the second question using a *cumsum chart* (Hawkins and Olwell, 1998), a tool that has also already appeared in the present context (Ellaway, 1978; Ushiba et al., 2002; Tam, 2009); but we bring *two decisive improvements*. Working with the difference of two observed PSTH *makes our test free*. The combination of Donsker's theorem (Billingsley, 1999; Durrett, 2009) with the Brownian confidence domain (Kendall et al., 2007) provides the first global test for an absence of drift in a cumsum chart context. Our simulation study shows that when 250 of more bins are used (a realistic number in practice) the actual test coverage probabilities are very close to the nominal ones for the most commonly used values (Table 2). They also show how to set the nominal value in order to get the right target value for as few as 50 bins.

From a neurophysiological perspective, the construction of confidence envelops—instead of the pointwise confidence intervals leads to a much safer eval-

uation of the presence / absence of a response for the class of neurons illustrated on Fig. 1, 2 and 4. The members of this class have a "high" spontaneous discharge rate (20-30 Hz) with "bursty" patterns (periods of relative silence alternate with periods of high discharge frequency). They tend to exhibit small responses when they do respond and this response / no response issue was not settled with our pointwise intervals. The members of the other, more frequent, neuronal classes have a more regular spontaneous discharge with a lower rate (5-10 Hz)—a member of this class is illustrated on Fig. 3—, they usually exhibit much more pronounced responses. A general comparison of the PSTH plus pointwise confidence intervals (Fig. 8, Pouzat and Chaffiol 2009) with the same PSTH plus a confidence envelop (Supplementary Files) makes this point. The same goes for the identity / non-identity of two responses to two different stimuli of a given neuron. For the neuron used on Fig. 3, the pointwise confidence intervals always overlap and even if they did not at best a tentative conclusion could be drawn, *while our new identity test gives us a very clear answer*. Our proposed methods drastically improve the interpretations of PSTH with a minimal investment since we provide implementations in both `R` and `Python`.

References

F. J. Anscombe. The transformation of poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4):246–254, 1948.

Patrick Billingsley. *Convergence of Probability Measures*. Wiley - Interscience, second edition, 1999.

D. R. Brillinger, H. L. Bryant, and J. P. Segundo. Identification of synaptic interactions. *Biol Cybern*, 22(4):213–228, 1976.

David R. Brillinger. Nerve Cell Spike Train Data Analysis: A Progression of Technique. *Journal of the American Statistical Association*, 87(418):260–271, June 1992.

Lawrence D. Brown, T. Tony Cai, and Anirban DasGupta. Interval estimation for a binomial proportion. *Statistical Science*, 16(2):101–117, 2001.

Lawrence D. Brown, T. Tony Cai, and Harrison H. Zhou. Nonparametric regression in exponential families. *The Annals of Statistics*, 38(4):2005–2046, 08 2010.

T. Brown. A martingale approach to the poisson convergence of simple point processes. *The Annals of Probability*, 6(4):615–628, 1978.

D. R. Cox and P. A. W. Lewis. *The Statistical Analysis of Series of Events*. John Wiley & Sons, 1966.

G.H. Dörrscheidt. The statistical significance of the peristimulus time histogram (psth). *Brain Research*, 220(2):397 – 401, 1981.

J. Durbin. Some methods of constructing exact tests. *Biometrika*, 48(1/2): 41–55, 1961.

Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 4th edition, 2009.

Jos J. Eggermont. *The correlative brain: theory and experiment in neural interaction*. Studies of brain function. Springer-Verlag, 1990.

Gaute T Einevoll, Felix Franke, Espen Hagen, Christophe Pouzat, and Kenneth D Harris. Towards reliable spike-train recordings from thousands of neurons with multielectrodes. *Current Opinion in Neurobiology*, 22(1):11–17, 2012.

P.H Ellaway. Cumulative sum technique and its application to the analysis of peristimulus time histograms. *Electroencephalography and Clinical Neurophysiology*, 45(2):302 – 304, 1978.

Murray F. Freeman and John W. Tukey. Transformations related to the angular and the square root. *The Annals of Mathematical Statistics*, 21(4):607–611, 1950.

George L. Gerstein and Nelson Y.-S. Kiang. An Approach to the Quantitative Analysis of Electrophysiological Data from Single Neurons. *Biophysical Journal*, 1(1):15–28, September 1960.

Edmund M. Glaser and Daniel S. Ruchkin. *Principles of neurobiological signal analysis*. Academic Press. Academic Press, 1976.

B. Grigelionis. On the convergence of sums of random step processes to a poisson process. *Theory of Probability & Its Applications*, 8(2):177–182, 1963.

D.M. Hawkins and D.H. Olwell. *Cumulative Sum Charts and Charting for Quality Improvement*. Information Science and Statistics. Springer New York, 1998.

Cari G. Kaufman, Valérie Ventura, and Robert E. Kass. Spline-based non-parametric regression for periodic functions and its application to directional tuning of neurons. *Statistics in Medicine*, 24(14):2255–2265, 2005.

W.S. Kendall, J.M. Marin, and C.P. Robert. Brownian Confidence Bands on Monte Carlo Output. *Statistics and Computing*, 17(1):1–10, 2007.

Peter A. W. Lewis. Some results on tests for poisson processes. *Biometrika*, 52 (1/2):67–77, 1965.

C. R. Loader and J. J. Deely. Computations of boundary crossing probabilities for the wiener process. *Journal of Statistical Computation and Simulation*, 27(2):95–105, 1987.

D. H. Perkel, G. L. Gerstein, and G. P. Moore. Neuronal spike trains and stochastic point processes. i the single spike train. *Biophys. J.*, 7:391–418, 1967.

Christophe Pouzat and Antoine Chaffiol. Automatic Spike Train Analysis and Report Generation. An Implementation with R, R2HTML and STAR. *J Neurosci Methods*, 181:119–144, 2009.

Christophe Pouzat and Antoine Chaffiol. Data set from Pouzat and Chaffiol (2009) Journal of Neuroscience Methods 181:119., Jan 2015. DOI:10.5281/zenodo.14281.

Christophe Pouzat and Georgios Is. Detorakis. Spysort: Neuronal spike sorting with python. In Pierre de Buyl and Nelle Varoquaux, editors, *Proceedings of the 7th European Conference on Python in Science (EuroSciPy 2014)*, pages 27 – 34, 2014.

Patricia Reynaud-Bouret, Vincent Rivoirard, Franck Grammont, and Christine Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *Journal of Mathematical Neuroscience*, 4:3, 2014.

H. Shimazaki and S. Shinomoto. A method for selecting the bin size of a time histogram. *Neural Computation*, 19(6):1503–1527, 2007.

H. Shimazaki and S. Shinomoto. Kernel bandwidth optimization in spike rate estimation. *Journal of Computational Neuroscience*, 2010.

David Siegmund. *Sequential analysis: tests and confidence intervals*. Springer, 1985.

Jiayang Sun and Clive R. Loader. Simultaneous confidence bands for linear regression and smoothing. *The Annals of Statistics*, 22(3):1328–1345, 1994.

David Tam. A theoretical analysis of cumulative sum slope (cusum-slope) statistic for detecting signal onset (begin) and offset (end) trends from back-ground noise level. *The Open Statistics and Probability Journal*, 1:43–51, 2009.

Junichi Ushiba, Yutaka Tomita, Yoshihisa Masakado, and Yoko Komune. A cumulative sum test for a peri-stimulus time histogram using the monte carlo method. *Journal of Neuroscience Methods*, 118(2):207 – 214, 2002.

Valérie Ventura, Roberto Carta, Robert E. Kass, Sonya N. Gettner, and Carl R. Olson. Statistical analysis of temporal evolution in single-neuron firing rates. *Biostat*, 3(1):1–20, 2002.

Patrick D. Wall. Repetitive discharge of neurons. *Journal of Neurophysiology*, 22(3):305–320, 1959.

Garrick Wallstrom, Jeffrey Liebner, and Robert E. Kass. An Implementation of Bayesian Adaptive Regression Splines (BARS) in C with S and R Wrappers. *Journal of Statistical Software*, 26(1):1–21, 2 2007.

Larry Wasserman. *All of Nonparametric Statistics*. Springer, 2006.

A Raster plots of the analyzed data

The usual way of presenting "raw" spike train data when a stimulus has been repetitively applied is the *raster plot* also referred to as a *dot display* (Wall, 1959; Glaser and Ruchkin, 1976; Eggermont, 1990)—Brillinger (1992) uses the term *rastor plot*—: neuronal responses are displayed one above the other aligned on the stimulus onset; individual spikes are represented by a dot, a tick or any other suitable glyph. Fig. 5 shows 15 successive responses to a 0.5 s air puff of citronellal of neuron 2 from experiment `e070528`. Fig. 6 shows 20 successive responses of neuron 1 of experiment `e060817` to 2 different odors: citronellal and terpineol.

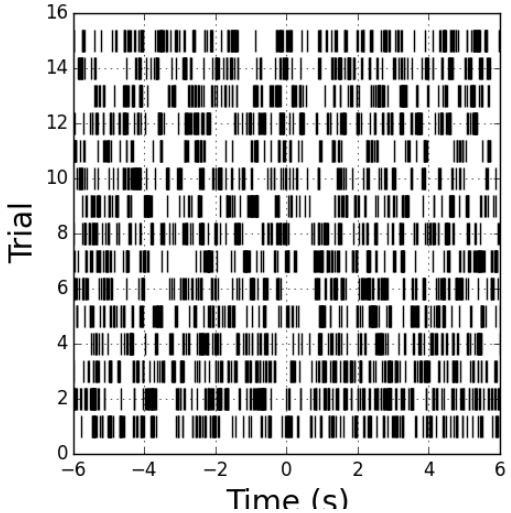

Figure 5: Example of 15 stimulations with citronellal for a neuron in the first olfactory relay of an insect (neuron 2 from experiment `e070528`). Each tick represents a spike occurrence. The first stimulation is at the bottom and the 15th at the top. Stimulations are delivered during 500 ms starting at time 0—in fact these 500 ms are the opening time of the valve diverting the continuous flux of moist air through a vial saturated with the odor; the odor presentation on the insect antenna does not stop when the valve closes but approximately 500 ms later—.

B Smooth stabilized PSTH of the terpineol and citronellal responses of neuron 1

In order to facilitate the interpration of Fig. 3 the smooth stabilized PSTH of the terpineol and citronellal responses of neuron 1 from experiment `e060817` are shown on Fig. 7.

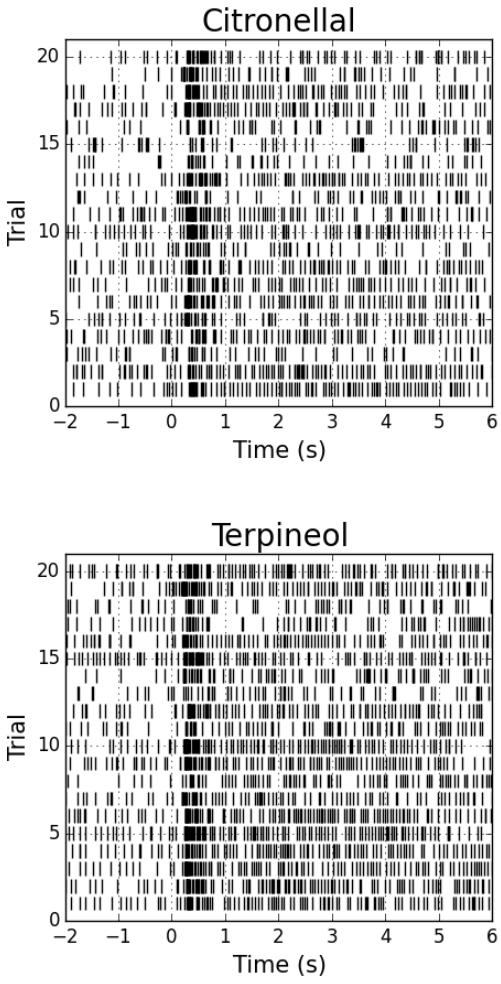

Figure 6: Neuron 1 from experiment `e060817`: 20 stimulations with citronellal and terpineol. The stimulus starts at time 0 (the valve was opened for 500 ms like in Fig. 5).

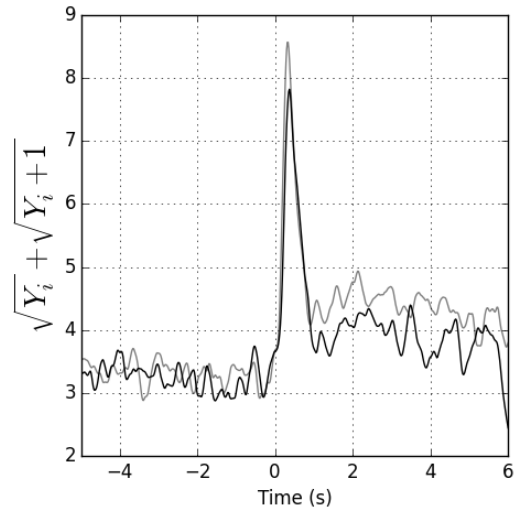

Figure 7: `SmoothStabilizedPSTH` instances computed from the citronellal (black) and terpineol (grey) responses of neuron 1 from experiment `e060817`.